

**USING PRINCIPAL COMPONENTS ANALYSIS AND CORRESPONDENCE
ANALYSIS FOR ESTIMATION IN LATENT VARIABLE MODELS**

BU-1461-M

November, 1999

Henry S. Lynn
Rho, Inc.
121 S. Estes Dr., Suite 100
Chapel Hill, NC 27514, USA

and

Charles E. McCulloch
Department of Biometrics
Cornell University
Ithaca, NY 14853, USA

Keywords: *correspondence analysis; principal component analysis; incidental parameters; consistency.*

This work was partially funded by NSF DMS-9625476

Abstract

Correspondence analysis (CA) and principal component analysis (PCA) are often used to describe multivariate data. In certain applications, they have been used for estimation in latent variable models. The theoretical basis for such inference is assessed in generalized linear models where the linear predictor equals $\alpha_j + x_i\beta_j$ or $a_j - b_j(x_i - u_j)^2$, ($i=1,\dots,n; j=1,\dots,m$), and x_i is treated as a latent fixed effect. The PCA and CA eigenvectors/column scores are evaluated as estimators of β_j and u_j and as estimators of u_j respectively. With m fixed and $n \uparrow \infty$, consistent estimators cannot be obtained due to the incidental parameters problem unless sufficient “moment” conditions are imposed on x_i . PCA is equivalent to maximum likelihood estimation for the linear Gaussian model, and gives a consistent estimator of β_j (up to a scale change) when the second sample moment of x_i is positive and finite in the limit. It is inconsistent for Poisson and Bernoulli distributions, but when b_j is constant its first and/or second eigenvectors can consistently estimate u_j (up to a location and scale change) for the quadratic Gaussian model. In contrast, the CA estimator is always inconsistent. For finite samples, however, the CA column scores often have high correlations with the u_j 's, especially when the response curves are spread out relative to one another. The correlations obtained from PCA are usually weaker. Although, the second PCA eigenvector can sometimes do much better than the first eigenvector, and for incidence data with tightly clustered response curves its performance is comparable to that of CA. For small sample sizes, PCA and particularly CA are competitive alternatives to maximum likelihood, and may be preferred because of their computational ease.

Keywords: correspondence analysis; principal component analysis; incidental parameters; consistency.

1. Introduction

Given data on m different characteristics for each of n objects, a researcher may choose to use dimension reduction techniques like principal component analysis (PCA) or correspondence analysis (CA) to graphically summarize and describe the $n \times m$ data matrix. On the other hand, models can be postulated and fitted to the data. For example, consider an ecological study where the data represent counts of 12 species of hunting spiders caught in 28 pitfall traps in a Dutch dune area (Ter Braak 1985, 1986). CA was used to generate species

scores that describe the ordering/arrangement of the different species, but Poisson regression models were also fitted to each species with respect to a latent environmental variable. The use of latent variable models is also pervasive in many disciplines like economics, sociology and psychology, and such models often allow researchers to better understand the underlying mechanisms that may have produced the data. Several authors have compared PCA and CA to least squares and ML estimation in latent variable models (Gauch *et al.* 1974, 1977, Goodall and Johnson 1982, Ihm and van Groenewoud 1984), but did not elaborate on the properties of the estimators. It is also widely known that PCA is equivalent to maximum likelihood (ML) estimation for a linear Gaussian model, and there is suggestion that CA may approximate ML estimation for generalized linear models (Ter Braak 1985). This paper addresses the theoretical basis for using PCA and CA for inference. Specifically, it examines the PCA and CA column scores as estimators of the parameters in latent variable generalized linear models. The column scores are chosen here since the row scores only supply information on the latent variable, which is not required to define the functional form of the predicted model.

Two points are important when examining PCA and CA as inferential methods. First, the proposed models treat the latent variable as a fixed effect with n levels to be estimated along with the other parameters. This limits inference to the designated objects in the data, but more importantly inference may be invalid when applying large sample approximations for ML estimators. Under the scenario where m is fixed but $n \uparrow \infty$, the ML estimator is inconsistent because of the incidental parameters problem (Neyman and Scott 1948), which is well-documented in the literature on functional measurement error models (Fuller 1987) and item response theory models (Wright 1977, Haberman 1977, De Leeuw and Verhelst 1986). For this reason, we focus on evaluating the consistency of the PCA and CA estimators, and Section 5 shows that sufficient “moment” conditions must be imposed on the latent variable before consistency can be ascertained. Intuitively, this suggests that we regard the latent variable as being from some population with finite moments, which is reminiscent of the

nonparametric ML approach (Keifer and Wolfowitz 1956, Follmann 1988). Second, the PCA and CA scores are only unique up to a scaling factor. Moreover, the models are shown in Section 2.1 to be overparameterized, and location and scale identifiability constraints are needed to obtain unique estimates of the parameters. Therefore, in assessing asymptotic behavior we only examine whether the PCA and CA column scores can consistently estimate the parameters up to a scale and location change (i.e. whether they have perfect correlation with the parameter values). The latter interpretation also suggests that it would be useful to compare results in terms of the correlation since it is a familiar location and scale invariant statistic of the estimates. Likewise, in the simulations we compare the finite sample performances of the PCA, CA, and ML estimates in terms of their correlation with the true parameter values.

2. Models

We consider two classes of generalized linear models where the linear predictor is either a linear or quadratic function of the latent variable. The first model assumes that the responses, y_{ij} ($i=1, \dots, n; j=1, \dots, m$), are modeled with

$$\eta_{ij} = \text{link}(\mu_{ij}) = \alpha_j + x_i \beta_j, \quad (1)$$

where $\mu_{ij} = E(y_{ij})$ and x_i is the latent fixed effect. The focus is on the Gaussian, Poisson and Bernoulli distributions, and canonical links are assumed in each case; i.e. the identity, log, and logit links respectively. For the Gaussian case, model (1) is sometimes called the functional factor analysis model (Anderson and Ameniya 1988), which is in essence a functional measurement error model. While for the Bernoulli case, it corresponds to the 2-parameter Rasch model in psychometric item response theory (Andersen 1980).

As an extension to model (1), consider η_{ij} as a quadratic function of x_i by letting

$$\eta_{ij} = a_j - \frac{1}{2} \frac{(x_i - u_j)^2}{t_j^2}. \quad (2)$$

This parameterization is popular in ecological applications, where different species of flora or fauna often occupy different niches in their habitat and their abundances thus vary unimodally with respect to some environmental gradient. a_j is the maximum of the expected response curve, t_j is the tolerance which measures the curve width, and u_j is the optimum or mode of the response curve. In this context, the latent variable x_i can be some hypothetical gradient or it can be some environmental variable that is too expensive to measure or can no longer be measured as in paleontological studies.

These two models are pertinent for a variety of applications. As stated above, model (1) encompasses the functional factor analysis model and the 2-parameter Rasch model, which are common in the social sciences. For example, Dawkins (1989) analyzed the completion times of various track races for 55 countries. The assumption of fixed country effects seems tenable, and as a first approximation one may expect the completion times to vary linearly with some latent athletic ability variable. In item response theory, treating x_i as fixed effects in the estimation dates back to the 1960s (Birnbaum 1968). Model (1) is also applicable to many industrial experiments and bioassays where reagent or dose concentrations are often considered as fixed effects. For example, a dose-response experiment may have been conducted at different titers of AZT drug concentrations, but the investigator may also be interested in modeling the relationship with the actual latent viremia levels.

When nonlinearity is expected, the quadratic model (2) is more appropriate. Ecologists regularly use the Poisson and Bernoulli cases to model species abundances in counts or incidences (Ter Braak 1987). The same model is also used for seriation in archaeology (Kendall 1971), a process by which artifacts excavated at different sites are typed and chronologically ordered. Beyond ecology and archaeology, there are a host of applications for model (2). For example, quadratic dose-response relationships are not uncommon (Cox 1972). An economist may model different types of consumption for different socio-geographic groups using model (2), with the socio-geographic groups being the fixed effects of interest. A latent living standard index can be hypothesized for the socio-

geographic groups, and unimodal response functions for each consumption type with respect to this latent index is justifiable since different groups may consume more goods of a particular type (Ihm and van Groenewoud 1984). In the medical setting, different hospitals may excel at different therapies; in a national election poll, different cities may favor different political parties (Van der Heijden *et al.* 1994); and finally in agricultural experiments, the crop yields for different treatment combinations may vary with soil type (Goodman and Haberman 1990).

In this paper, the latent effects x_i are treated as nuisance parameters since the objective is to recover the underlying model by estimating α_j and β_j in model (1) or a_j , u_j , and t_j in model (2). However, these parameters are intrinsically aliased (McCullagh and Nelder 1989) and identifiability constraints must be imposed to obtain unique estimates.

2.1 Identifiability

To illustrate what constraints are required, (1) and (2) are examined as a series of simpler models. Consider first model (1) and suppose initially that

$$\eta_{ij} = x_i \beta_j.$$

If x_i, β_j is a solution then $x_i^* = x_i/\sigma$, $\beta_j^* = \sigma\beta_j$ is also a solution, where $\sigma \in \mathfrak{R}$ but nonzero. One restriction is thus required to obtain unique estimates; e.g. $\hat{x}_1=1$ or $\hat{\beta}_1=1$. (Note however that exceptions may arise for certain data configurations.) With the full model $\eta_{ij} = \alpha_j + x_i \beta_j$, α_j and β_j are aliased. Even if all the α_j 's or all the β_j 's are known, the other parameters are still not identifiable. The crucial parameter is x_i , and if there are two constraints on \hat{x}_i (e.g. $\hat{x}_1=1$ and $\hat{x}_2=2$) then unique estimates can be obtained for the rest. In general, two restrictions (e.g. $\hat{x}_1=1$, $\hat{\beta}_1=1$) are required, and one of which must be on the \hat{x}_i 's.

Next consider model (2) and start with

$$\eta_{ij} = (x_i - u_j)^2.$$

If x_i, u_j is a solution then $x_i^* = \pm(x_i + \Delta)$, $u_j^* = \pm(u_j + \Delta)$ is also a solution, where $\Delta \in \mathfrak{R}$. Hence two restrictions are needed; e.g. $\hat{x}_1=1$ and $\hat{u}_1=1$. Now suppose

$$\eta_{ij} = a_j - (x_i - u_j)^2, \quad (3)$$

which like model (1) also has three sets of parameters, a_j , x_i , and u_j . For model (1) it is sufficient to have two constraints, and in general this is also true for (3) except that n must be greater than 2; otherwise there would be more parameters than equations for each j . For the full model (2), a_j and b_j are aliased, which suggests that one more constraint is required. Indeed, if \hat{x}_1 and \hat{u}_1 are set a priori then setting \hat{a}_1 would uniquely determine \hat{b}_1 , the other \hat{x}_i 's, and all the subsequent estimates. In general, three constraints are sufficient to ensure identifiability for model (2), with one constraint each on \hat{x}_i and \hat{u}_j ; e.g. $\hat{a}_1=1$, $\hat{u}_2=1$, $\hat{x}_3=1$.

The above constraints are merely rules for identifying the estimates and are not part of the model. However, they do imply that two researchers may get different estimates when analyzing the same data since there are an infinite number of possible constraints. One option to resolve this indeterminacy is to use quantities that are invariant with respect to the type of constraints. For example, the scale invariant statistics of \hat{x}_i and $\hat{\beta}_j$ are estimable in model (1), and the scale and location invariant statistics of \hat{x}_i and \hat{u}_j are estimable in model (2). Likewise, we compare the PCA, CA, and ML estimates of u_j using the correlation coefficient since it is one of the more popular scale and location invariant statistics.

3. Principal Component Analysis

PCA provides an orthogonal least squares approximation to the data via a singular value decomposition. We focus on mean-centered PCA since this formulation is most often compared to CA. Formally, if $\mathbf{W} \equiv \{y_{ij} - \bar{y}_{.j}\}$ has rank $R \in \{1, \dots, m\}$, the singular value decomposition of \mathbf{W} gives

$$\mathbf{W} = \sum_{r=1}^R \lambda_r \mathbf{p}_r \mathbf{q}_r', \text{ such that} \quad (4)$$

$$\mathbf{W}'\mathbf{W}\mathbf{q}_r = \lambda_r^2 \mathbf{q}_r, \quad (5)$$

$$\mathbf{W}\mathbf{W}'\mathbf{p}_r = \lambda_r^2 \mathbf{p}_r, \quad (6)$$

$$\mathbf{p}_r' \mathbf{p}_l = \mathbf{q}_r' \mathbf{q}_l = \delta_{rl},$$

where δ_{rl} is Kronecker's delta, and \mathbf{p}_r , \mathbf{q}_r define the row and column scores respectively. Equation (5) is the usual definition of PCA, where \mathbf{q}_r is the r th principal component given by the eigenvector of the covariance matrix $\mathbf{W}'\mathbf{W}$. The covariance, rather than the correlation, matrix is used since for the applications considered in this paper the m characteristics represent the same type of variable and are measured using the same units; e.g. the completion times of m different races or the presence of particular traits in m artifacts. Analysis of the covariance matrix also makes interpretation of quantities like “variance explained” and hypothesis testing more straightforward, and is equivalent to Euclidean distance multidimensional scaling.

3.1 Relationship with Maximum Likelihood

PCA is usually used for exploratory descriptive purposes, and does not explicitly require a model. However, PCA can also be viewed as multivariate linear regression with latent predictors. Consider model (1) with Gaussian errors, the log-likelihood is then proportional to

$$-mn\log(2\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{i=1}^n \text{tr}((\mathbf{y}_i - \boldsymbol{\alpha} - x_i\boldsymbol{\beta})(\mathbf{y}_i - \boldsymbol{\alpha} - x_i\boldsymbol{\beta})'),$$

where $\mathbf{y}_i = (y_{i1}, \dots, y_{im})'$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)'$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)'$. The likelihood equations for β_j and x_i do not involve α_j and σ^2 , and can be shown to respectively equal

$$c_1\boldsymbol{\beta} = \mathbf{W}'\mathbf{x} \text{ and} \tag{7a}$$

$$c_2\mathbf{x} = \mathbf{W}\boldsymbol{\beta}, \tag{7b}$$

with $c_1 = \sum_i (x_i - \bar{x})^2$, $c_2 = \sum_j \beta_j^2$, $\mathbf{x} = (x_1 - \bar{x}, \dots, x_n - \bar{x})'$. Combining these two equations we have

$$\mathbf{W}'\mathbf{W}\boldsymbol{\beta} = c_1c_2\boldsymbol{\beta}, \text{ and} \tag{8}$$

$$\mathbf{W}\mathbf{W}'\mathbf{x} = c_1c_2\mathbf{x}, \tag{9}$$

which shows that $\lambda^2 = c_1c_2$ is an eigenvalue, and $\boldsymbol{\beta}$ and \mathbf{x} are eigenvectors. In fact, substituting the expressions for the ML estimators of α_j , x_i , β_j , σ^2 into the log-likelihood, we find that the log-likelihood varies inversely according to

$$tr((\mathbf{W} - \mathbf{x}\beta')'(\mathbf{W} - \mathbf{x}\beta')) = tr(\mathbf{W}'\mathbf{W}) + tr(\beta'\beta\mathbf{x}'\mathbf{x}) - 2tr(\beta\mathbf{x}'\mathbf{W}) = tr(\mathbf{W}'\mathbf{W}) - \lambda^2.$$

The log-likelihood attains its maximum when λ^2 is largest, implying that the dominant eigenvectors of (8) and (9) are also the ML estimators. Now (8) and (9) have the same form as (5) and (6), when a rank one approximation is applied to \mathbf{W} . Therefore, solving for p_i and q_j in PCA is equivalent to calculating the ML estimates of $x_i - \bar{x}$ and β_j in model (1), although unlike ML, PCA does not directly estimate the α_j and σ^2 .

The fact that the ML estimators for β and \mathbf{x} are eigenvectors suggests that they can be computed using standard algorithms for calculating eigenvectors without requiring joint maximization via Newton-Raphson. Moreover, equations (7a) and (7b) show that β is obtained by regressing the rows of \mathbf{W} on \mathbf{x} , whereas \mathbf{x} is obtained by regressing the columns of \mathbf{W} on β . The ML estimates can thus be calculated by simply iterating between equations (7a) and (7b) until convergence, which is an example of the Gauss-Seidel-Newton method (Thisted 1988). This algorithm also provides a general framework for computing ML estimates in model (1) by iteratively performing “regressions” (i.e. solving for α_j, β_j by fixing the x_i ’s to some initial values) and “calibrations” (i.e. solving for x_i using the intermediate estimates of α_j, β_j).

4. Correspondence Analysis

CA is a multi-faceted technique dating back to the 1930s. It is mathematically equivalent to canonical correlation analysis, dual-scaling, and reciprocal averaging, with applications in biometry, psychology, and ecology respectively (Greenacre 1984). Like PCA, CA applies a singular value decomposition to the data, except the data are weighted inversely by the square root of the row and column sums.

Let $\mathbf{C} = \text{diag}(y_{.j})$ be a $m \times m$ diagonal matrix with $y_{.j} = \sum_i y_{ij} > 0$, and $\mathbf{R} = \text{diag}(y_{i.})$ a $n \times n$ diagonal matrix with $y_{i.} = \sum_j y_{ij} > 0$. CA can be defined by

$$\mathbf{R}^{-1/2} \mathbf{Y} \mathbf{C}^{-1/2} = \sum_{r=1}^R \lambda_r \mathbf{p}_r \mathbf{q}_r', \text{ such that} \quad (10)$$

$$\mathbf{R}^{-1}\mathbf{Y}\mathbf{C}^{-1}\mathbf{Y}'\mathbf{z}_r = \lambda_r^2 \mathbf{z}_r, \text{ and} \quad (11)$$

$$\mathbf{C}^{-1}\mathbf{Y}'\mathbf{R}^{-1}\mathbf{Y}\mathbf{v}_r = \lambda_r^2 \mathbf{v}_r, \quad (12)$$

where $\mathbf{z}_r = \lambda_r \mathbf{R}^{-1/2} \mathbf{p}_r$, $\mathbf{v}_r = \mathbf{C}^{-1/2} \mathbf{q}_r$ are the row and column scores respectively. Using (10), the scores can be combined to give

$$\mathbf{z}_r = \mathbf{R}^{-1}\mathbf{Y}\mathbf{v}_r \text{ and } \lambda_r^2 \mathbf{v}_r = \mathbf{C}^{-1}\mathbf{Y}'\mathbf{z}_r, \text{ or}$$

$$z_{ri} = \frac{\sum_j y_{ij} v_{rj}}{\sum_j y_{ij}} \text{ and } \lambda_r^2 v_{rj} = \frac{\sum_i y_{ij} z_{ri}}{\sum_i y_{ij}}, \quad (13)$$

which shows that row scores are weighted averages of the column scores and vice versa, hence the term “reciprocal averaging”. It is easily verified that a solution to (13) is $\lambda_1=1$ with the corresponding row and column scores all equal to 1, and λ_1 is also the largest singular value (Hill 1974). This trivial or zero order solution is generally ignored, and only higher order solutions are retained.

4.1 Approximations

Since PCA is equivalent to ML estimation for the Gaussian case in model (1), it is natural to ask whether CA might also have any model-based relationship. Researchers have found that CA provides reasonable descriptions of the data when responses are unimodal, and this may be attributed to several explanations. Lancaster (1957) showed that when two variables X_1 and X_2 have a bivariate normal density, the CA scores correspond to Tchebycheff-Hermite polynomials of X_1 and X_2 : the first order CA solutions are first order Hermite polynomials (which are X_1 and X_2), and the second order CA solutions are second order Hermite polynomials (which are proportional to X_1^2 and X_2^2). For the Poisson case in model (2), μ_{ij} has a similar form as the bivariate normal density since $\mu_{ij} \propto \exp(-(x_i - u_j)^2)$ even though x_i and u_j are not random variables. Hence, the CA scores may approximate the Hermite polynomials of x_i and u_j . This also explains why a plot of the first and second order CA solutions can exhibit an arch shape, which is sometimes called the horseshoe effect (Hill 1973).

There are also similarities between CA and ML estimation for model (2). Ter Braak (1985) indicated that likelihood equations for x_i and u_j can be rearranged as:

$$x_i = \sum_j \frac{y_{ij} u_j}{t_j^2} \bigg/ \sum_j \frac{y_{ij}}{t_j^2} + \left[\sum_j \frac{(x_i - u_j) \mu_{ij}}{t_j^2} \bigg/ \sum_j \frac{y_{ij}}{t_j^2} \right] \text{ and}$$

$$u_j = \sum_i \frac{y_{ij} x_i}{y_{.j}} - \left[\sum_i \frac{(x_i - u_j) \mu_{ij}}{y_{.j}} \right].$$

When t_j is constant and *if* the terms in the square brackets are negligible, these reduce to the equations in (13) for defining the CA row and column scores respectively.

5. Consistency of Estimators

In order to formally compare PCA and CA, we assess the asymptotic performances of the estimators by checking whether they can consistently estimate the parameters up to a scale and location change. The underlying asymptotic theory assumes m to be fixed and $n \uparrow \infty$, which seems appropriate for the applications cited in Section 2 where there are a finite number of characteristics (e.g. species of spiders or types of artifacts) but conceivably a much larger sampling frame for objects (e.g. sites or graves). We begin with PCA by examining the asymptotic behavior of its column scores as estimators of β in model (1) and as estimators of \mathbf{u} in model (2). Next, the CA column scores are examined as estimators of \mathbf{u} . For both the PCA and CA estimators, consistency is first examined theoretically via eigenvector solutions, and the results are then verified numerically.

5.1 PCA Estimator

The PCA estimator is taken to be the dominant eigenvector $\hat{\mathbf{w}}$ associated with the largest eigenvalue, λ_{\max} , of $\mathbf{W}'\mathbf{W}$ defined in (6) with (j,k) th element equal to

$$\sum_i y_{ij} y_{ik} - \frac{1}{n} \sum_i y_{ij} \sum_i y_{ik}, \quad (j,k=1,\dots,m). \quad (14)$$

To examine the asymptotic property of $\hat{\mathbf{w}}$, we first establish the conditions that are required for $n^{-1}\mathbf{W}'\mathbf{W}$ to converge to Σ :

a) If the y_{ij} 's are independent Gaussian random variables with mean μ_{ij} and variance σ_e^2 , then

$$Var(\sum_i y_{ij}) = n\sigma_e^2 = o(n^2), \text{ and}$$

$$Var(\sum_i y_{ij}y_{ik}) = \sum_i \{\mu_{ij}^2 Var(y_{ik}) + \mu_{ik}^2 Var(y_{ij}) + Var(y_{ij})Var(y_{ik})\}.$$

For model (1) $\sum_i \mu_{ij}^2 = n\alpha_j^2 + 2\alpha_j\beta_j \sum_i x_i + \beta_j^2 \sum_i x_i^2$. When

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_i x_i^2 = h, \text{ where } h \in (0, \infty),$$

$\sum_i \mu_{ij}^2 = o(n^2)$ and likewise $Var(\sum_i y_{ij}y_{ik}) = o(n^2)$. For model (2), $\mu_{ij} = a_j - b_j(x_i - u_j)^2 \leq a_j$ (since $b_j > 0$) and thus $Var(\sum_i y_{ij}y_{ik}) = o(n^2)$ again.

b) If the y_{ij} 's are independent Poisson counts, then

$$Var(\sum_i y_{ij}) = \sum_i \mu_{ij} \text{ and } Var(\sum_i y_{ij}y_{ik}) = \sum_i (\mu_{ij}^3 + \mu_{ik}^3 + \mu_{ij}\mu_{ik}).$$

Suppose $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_i \exp(x_i^2)$ is positive and finite, then under model (1) $\sum_i \mu_{ij}$ and $Var(\sum_i y_{ij}y_{ik})$ are both $o(n^2)$. For model (2) $\mu_{ij} \leq \exp(a_j)$, and $Var(\sum_i y_{ij})$ and $Var(\sum_i y_{ij}y_{ik})$ are again both $o(n^2)$. Lastly,

c) if the y_{ij} 's are independent Bernoulli variables then y_{ij} and $y_{ij}y_{ik}$ are both uniformly bounded.

Therefore the conditions for the Weak Law of Large Numbers are met for all three distributions in models (1) and (2), implying that

$$\frac{1}{n} \sum_i y_{ij} \rightarrow \frac{1}{n} \sum_i E(y_{ij}) \text{ and } \frac{1}{n} \sum_i y_{ij}y_{ik} \rightarrow \frac{1}{n} \sum_i E(y_{ij}y_{ik}).$$

It follows from (14) that $n^{-1}\mathbf{W}'\mathbf{W} \rightarrow \mathbf{\Sigma}$, which has (j,k) th element equal to

$$\frac{1}{n} \sum_i (Var(y_{ij})\delta_{jk} + \mu_{ij}\mu_{ik}) - \frac{1}{n} \sum_i \mu_{ij} \frac{1}{n} \sum_i \mu_{ik}. \quad (15)$$

When λ_{\max} of $\mathbf{W}'\mathbf{W}$ is unique (i.e. has multiplicity one), $\hat{\mathbf{w}}$ is a continuous function of the elements of $\mathbf{W}'\mathbf{W}$ (Ortega 1972, p.45). Let \mathbf{w} be the dominant eigenvector of $\mathbf{\Sigma}$ satisfying the equation $\lambda_{\mathbf{w}}\mathbf{w} = \mathbf{\Sigma}\mathbf{w}$. If $\lambda_{\mathbf{w}}$ is also unique, then $\hat{\mathbf{w}}$ converges to \mathbf{w} the dominant eigenvector of $\mathbf{\Sigma}$. We declare $\hat{\mathbf{w}}$ to be a consistent estimator of $\boldsymbol{\theta}$, which is either $\boldsymbol{\beta}$ in (1) or \mathbf{u} in (2), when \mathbf{w} satisfies the equation $\lambda\mathbf{w} = k_1\mathbf{1} + k_2\boldsymbol{\theta}$ or equivalently

$$\lambda w_j = k_1 + k_2 \theta_j, \quad (16)$$

where $k_1, k_2 \in \mathfrak{R}$, $k_2 \neq 0$, and $\mathbf{1}$ is a $m \times 1$ unit vector. This implies that asymptotically $\hat{\mathbf{w}}$ has perfect correlation with $\boldsymbol{\theta}$. Conversely, if $\lambda \mathbf{w} \neq k_1 \mathbf{1} + k_2 \boldsymbol{\theta}$ then $\hat{\mathbf{w}}$ is not a consistent estimator of $\boldsymbol{\theta}$.

Without loss of generality, consider the j th element of \mathbf{w} . Using (15), it can be shown after some algebraic simplifications that for the Gaussian case:

$$\lambda_w w_j = \sum_k w_k (\sigma_e^2 \delta_{jk} + \beta_j \beta_k \sigma_x^2) \quad (17)$$

in model (1), while for model (2)

$$\begin{aligned} \lambda_w w_j &= \sum_k w_k [\sigma_e^2 \delta_{jk} + b_j b_k (\gamma_{4x} - 2\gamma_{3x}(u_j + u_k) + 4\sigma_x^2 u_j u_k)] \\ &= \sigma_e^2 w_j + b_j (\gamma_{4x} \sum_k b_k w_k - 2\gamma_{3x} \sum_k b_k u_k w_k) + b_j u_j (4\sigma_x^2 \sum_k b_k u_k w_k - 2\gamma_{3x} \sum_k b_k w_k), \end{aligned} \quad (18)$$

where $\gamma_{4x} = \frac{\sum x_i^4}{n} - \frac{(\sum x_i^2)^2}{n^2}$, $\gamma_{3x} = \frac{\sum x_i^3}{n} - \frac{\sum x_i}{n} \frac{\sum x_i^2}{n}$, $\sigma_x^2 = \frac{\sum x_i^2}{n} - \frac{(\sum x_i)^2}{n^2}$.

For the Poisson case, we have for model (1)

$$\begin{aligned} \lambda w_j &= \frac{1}{n} \sum_i w_j \exp(\alpha_j + \beta_j x_i) + \frac{1}{n} \sum_i [\exp(\alpha_j + \beta_j x_i) \sum_k w_k \exp(\alpha_k + \beta_k x_i)] \\ &\quad - \frac{1}{n} \sum_i [\exp(\alpha_j + \beta_j x_i)] \frac{1}{n} \sum_i [\sum_k w_k \exp(\alpha_k + \beta_k x_i)], \end{aligned} \quad (19)$$

while for model (2)

$$\begin{aligned} \lambda w_j &= \frac{1}{n} \sum_i w_j \exp(a_j - b_j(x_i - u_j)^2) + \frac{1}{n} \sum_i [\exp(a_j - b_j(x_i - u_j)^2) \sum_k w_k \exp(a_k - b_k(x_i - u_k)^2)] \\ &\quad - \frac{1}{n} \sum_i [\exp(a_j - b_j(x_i - u_j)^2)] \frac{1}{n} \sum_i [\sum_k w_k \exp(a_k - b_k(x_i - u_k)^2)]. \end{aligned} \quad (20)$$

Finally for the Bernoulli case,

$$\begin{aligned} \lambda w_j &= \frac{1}{n} \sum_i w_j r_{ij} (1 + r_{ij})^{-2} + \frac{1}{n} \sum_i [(1 + r_{ij})^{-1} \sum_k w_k (1 + r_{ik})^{-1}] - \\ &\quad \frac{1}{n} \sum_i (1 + r_{ij})^{-1} \frac{1}{n} \sum_i [\sum_k w_k (1 + r_{ik})^{-1}], \end{aligned} \quad (21)$$

where $r_{ij} = \exp(-\alpha_j - \beta_j x_i)$ for model (1) and $r_{ij} = \exp(-a_j + b_j(x_i - u_j)^2)$ for model (2).

For the Poisson and Bernoulli distributions, the asymptotic eigenvector equations (19), (20) and (21) do not have the simple linear form as in (16), suggesting that $\hat{\mathbf{w}}$ is inconsistent.

However, for the Gaussian case the eigenvector equation (17) is linear in β_j and the following proposition holds.

Proposition 1. Define $S_n(x_i) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_i x_i^2$. If $S_n(x_i) \in (0, \infty)$, and $\lambda_{\max}, \lambda_w$ (the largest eigenvalues of $\mathbf{W}'\mathbf{W}$ and Σ respectively) are unique, then under model (1)

$$n^{-1} \hat{\mathbf{w}} \rightarrow (\sigma_e^2 + \sigma_x^2 \beta' \beta) \beta.$$

Proof. When $S_n(x_i)$ is positive and finite, $\sigma_x^2 \neq 0$ and $n^{-1} \mathbf{W}'\mathbf{W}$ converges to a well-defined Σ , which according to (17) equals

$$\sigma_e^2 \mathbf{I} + \sigma_x^2 \beta \beta'.$$

Now the $m \times m$ matrix $\sigma_x^2 \beta \beta'$ has rank 1, and it has $m-1$ eigenvalues with value zero and one non-trivial eigenvalue $\sigma_x^2 \beta' \beta$ (with corresponding eigenvector β). Since $\sigma_x^2 \beta \beta'$ and Σ share the same eigenvectors, this implies that Σ has $m-1$ eigenvalues with value σ_e^2 and largest eigenvalue $\sigma_e^2 + \sigma_x^2 \beta' \beta$, corresponding to its dominant eigenvector β .

Proposition 1 actually establishes when the ML estimator of β will be consistent since the PCA estimator is also the ML estimator for the Gaussian case in model (1). The crucial assumption is for $S_n(x_i)$, the second sample moment, to be positive and finite in the limit. This is in fact the univariate version of Gleser's (1981) assumption C for showing consistency of estimators in linear functional measurement error models. PCA, therefore, implicitly assumes a linear model, yet interestingly the PCA estimator can also be consistent for the Gaussian quadratic model in (2) when the tolerances are all equal.

Proposition 2. Let $S_n^r(x_i) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n |x_i^r|$. If $S_n^r(x_i) < \infty$ for $r=\{1,2,3,4\}$, $\sigma_x^2 \neq 0$, $b_j=b$ for all j , λ_{\max} and λ_w are unique, and $\lambda_w > \sigma_e^2$, then under model (2)

$$n^{-1} \hat{\mathbf{w}} \rightarrow k_1 \mathbf{1} + k_2 \mathbf{u},$$

where $k_1 = \gamma_{4x} \mathbf{1}' \mathbf{w} - 2\gamma_{3x} \mathbf{u}' \mathbf{w}$, $k_2 = 2\sigma_x^2 \mathbf{u}' \mathbf{w} - \gamma_{3x} \mathbf{1}' \mathbf{w}$, and $\gamma_{4x}, \gamma_{3x}, \sigma_x^2$ are defined in (18).

Proof. When $S_n^r(x_i)$ is finite, the terms γ_{4x}, γ_{3x} , and σ_x^2 are finite and $n^{-1} \mathbf{W}'\mathbf{W}$ converges to a well defined Σ . Furthermore, with $\sigma_x^2 \neq 0$ and $b_j=b$ for all j , (18) implies that

$$\Sigma = \sigma_e^2 \mathbf{I} + b^2 \mathbf{1} (\gamma_{4x} \mathbf{1}' - 2\gamma_{3x} \mathbf{u}') + b^2 \mathbf{u} (4\sigma_x^2 \mathbf{u}' - 2\gamma_{3x} \mathbf{1}'). \quad (22)$$

Let $\Sigma_1 = \Sigma - \sigma_e^2 \mathbf{I}$ and note that Σ_1 has rank 2. Thus it has $m-2$ trivial eigenvectors with zero as their eigenvalues, and according to (22) its two non-trivial eigenvectors are proportional to

$$\Sigma_1 \mathbf{w} = k_1 \mathbf{1} + k_2 \mathbf{u}, \text{ where } k_1 = \gamma_{4x} \mathbf{1}' \mathbf{w} - 2\gamma_{3x} \mathbf{u}' \mathbf{w}, k_2 = 2\sigma_x^2 \mathbf{u}' \mathbf{w} - \gamma_{3x} \mathbf{1}' \mathbf{w}.$$

It follows that Σ also has $m-2$ trivial eigenvectors with eigenvalue σ_e^2 , and when $\lambda_w > \sigma_e^2$ its dominant eigenvector corresponds to one of the two non-trivial eigenvectors of Σ_1 . Therefore $n^{-1} \hat{\mathbf{w}} \rightarrow k_1 \mathbf{1} + k_2 \mathbf{u}$, and the PCA estimator is a location and scale consistent estimator of \mathbf{u} .

Proposition 2 is algebraically similar to Kooijman's (1977) derivation of the eigenvector of $\mathbf{Z}'\mathbf{Z}$, where

$$\mathbf{Z} = \{z_{ij} - \bar{z}_{i.} - \bar{z}_{.j} + \bar{z}_{..}\} \text{ and } z_{ij} = a_j - b(x_i - u_j)^2,$$

although Kooijman assumed the data to be fixed rather than Gaussian random variables. In addition, it showed that there are two non-trivial eigenvectors of Σ , which are both linear in \mathbf{u} when their associated eigenvalues are greater than σ_e^2 . This implies that both the first and the second eigenvector of $\mathbf{W}'\mathbf{W}$ are consistent estimators of \mathbf{u} . However, when the x_i 's are symmetric about zero and the u_j 's sum to zero, then either the first or the second PCA eigenvector is consistent under the conditions of the following corollary.

Corollary. If $\gamma_{3x}=0$, $\mathbf{u}'\mathbf{1}=0$, and $m\gamma_{4x} < 4\sigma_x^2 \mathbf{u}'\mathbf{u}$, then $\hat{\mathbf{w}}$ is consistent up to a scale change.

Proof. When $\gamma_{3x}=0$, (22) implies that

$$\Sigma = \sigma_e^2 \mathbf{I} + b^2 \mathbf{1}(\gamma_{4x} \mathbf{1}') + b^2 \mathbf{u}(4\sigma_x^2 \mathbf{u}').$$

Therefore with $\mathbf{u}'\mathbf{1}=0$, the two non-trivial eigenvectors of Σ are $\mathbf{1}$ and \mathbf{u} with eigenvalues $mb^2\gamma_{4x} + \sigma_e^2$ and $4b^2\sigma_x^2 \mathbf{u}'\mathbf{u} + \sigma_e^2$ respectively.

The Corollary implies that if $m\gamma_{4x} > 4\sigma_x^2 \mathbf{u}'\mathbf{u}$ then the second (rather than the first) PCA eigenvector should be used to estimate \mathbf{u} . Likewise, both the first and the second PCA eigenvectors will be examined in the simulations since in certain situations one may be preferred over the other.

The eigenvector equations (19), (20) and (21) suggest that \mathbf{w} is not linear in θ , implying that $\hat{\mathbf{w}}$ is inconsistent. This can also be verified numerically by calculating \mathbf{w} for selected values of the parameters since \mathbf{w} is simply the eigenvector of Σ given in (15) and is a

function of μ_{ij} . However, the expression for Σ in (15) cannot be used directly since it depends on n , and we are interested in its numerical value as $n \uparrow \infty$. Therefore, in order to remove the dependency on n and to work out a specific counterexample, we assume $x_i \sim \text{iid } N(0,1)$ and apply a conditioning argument. This assumption also concurs with the “moment” conditions used on the x_i 's to establish consistency. The (j,k) th element Σ then equals

$$E_X[\text{Var}(y_{ij}|x_i)]\delta_{jk} + E_X[E(y_{ij}|x_i)E(y_{ik}|x_i)] - E_X[E(y_{ij}|x_i)]E_X[E(y_{ik}|x_i)], \quad (23)$$

where $E_X(\cdot)$ is the expectation with respect to the standard Gaussian density. It follows that for the Gaussian case (23) equals

$$\begin{aligned} &\sigma_e^2 \delta_{jk} + \beta_j \beta_k \text{ for model (1), and} \\ &\sigma_e^2 \delta_{jk} + 2b_j b_k (1 + 2u_j u_k) \text{ for model (2).} \end{aligned}$$

For Poisson counts, the moment generating function for non-central chi-square distributions can be used to show that (23) equals

$$\exp(\alpha_j + \beta_j^2/2)\delta_{jk} + \exp(\alpha_j + \alpha_k)\exp((\beta_j^2 + \beta_k^2)/2)(\exp(\beta_j \beta_k) - 1) \text{ for model (1).}$$

While for model (2) it equals

$$s_j \delta_{jk} + \exp(a_j - b_j u_j^2 + a_k - b_k u_k^2) \exp(d_{jk}/c_{jk})(1 - c_{jk})^{-1/2} - s_j s_k,$$

where $s_j = \exp(a_j) \exp(-b_j u_j^2 (1 + 2b_j)^{-1}) (1 - 2b_j)^{-1/2}$, $c_{jk} = 1 + 2(b_j + b_k)$, and

$$d_{jk} = -(b_j u_j^2 + b_k u_k^2) - 2b_j b_k (-u_j - u_k)^2.$$

In the Bernoulli case, there are no closed form solutions but the components of (23) can be calculated by numerical integration.

In order to calculate \mathbf{w} using the previous expressions for (23), we set $m=31$ and for illustrative purposes let $\alpha=E(0,0.3)$ and $\beta=E(0.5,0.8)$ in model (1), where $E(p,q)$ denotes a vector of equally spaced elements between (p,q) . Numerical integration was performed using Simpson's method with built-in functions in GAUSS (Aptech Systems 1992). Figure 1 displays the results for the Poisson and Bernoulli cases by plotting the residuals $w_j - \beta_j$ against the true values β_j . If \mathbf{w} is linear in β then the plot would show a straight line. Both curves are however quadratic, implying that the PCA estimator $\hat{\mathbf{w}}$ is inconsistent. For model (2), we set $\sigma_e^2=1$, $\mathbf{a}=E(30,32.25)$, $\mathbf{b}=E(1,4)$, and $\mathbf{u}=E(0,0.4)$ for the Gaussian case, and $\mathbf{a}=E(0.5,2)$,

$\mathbf{b}=E(1,2.5)$, and $\mathbf{u}=E(-2,2)$ for the Poisson and Bernoulli cases. Once again, the PCA asymptotic solutions produced curvilinear plots (Figures 2a, 2b and 2c) for all three distributions, confirming that $\hat{\mathbf{w}}$ is in general inconsistent for model (2).

5.2 CA Estimator

Next we inspect the asymptotic consistency of the CA column scores as estimators of \mathbf{u} in model (2). Note that $y_{i\cdot}$ and $y_{\cdot j}$ must be positive according to the definition of CA in (10), but there is a nonzero probability that $y_{i\cdot}$ and $y_{\cdot j}$ can be non-positive. Thus, in order to assess consistency, we assume that y_{ij} has expectation and variance according to model (2) but conditional on $y_{i\cdot}$ and $y_{\cdot j}$ being positive. This is still a reasonable working model since if any $y_{i\cdot}$ and $y_{\cdot j}$ is zero one would in practice simply reduce the dimensions of the data matrix. Using these assumptions, we suggest by analogy that the asymptotic solution of the CA estimator is not linear in \mathbf{u} by showing that its Taylor's series approximation cannot be written as $k_1\mathbf{1} + k_2\mathbf{u}$, $(k_1, k_2) \in \mathbb{R}$.

Let the CA estimator $\hat{\mathbf{v}}$ be the first order solution of the column scores in (12); i.e. the eigenvector associated with the second largest eigenvalue in (12). When the eigenvalue of $\hat{\mathbf{v}}$ is unique and $\mathbf{C}^{-1}\mathbf{Y}'\mathbf{R}^{-1}\mathbf{Y} \rightarrow \mathbf{\Sigma}$, $\hat{\mathbf{v}}$ converges to \mathbf{v} , the second eigenvector of $\mathbf{\Sigma}$. First, we establish the conditions for $\mathbf{C}^{-1}\mathbf{Y}'\mathbf{R}^{-1}\mathbf{Y}$ to converge to $\mathbf{\Sigma}$. The (j,k) th element of $\mathbf{C}^{-1}\mathbf{Y}'\mathbf{R}^{-1}\mathbf{Y}$ equals

$$\sum_i \frac{y_{ij}y_{ik}}{y_{i\cdot}y_{\cdot j}} = \frac{n}{\sum_i y_{ij}} \frac{1}{n} \sum_i \frac{y_{ij}y_{ik}}{y_{i\cdot}}.$$

Let $y_{i[m]}$ be the largest element in the series $y_{i\cdot}$. Since $y_{i\cdot}$ is by definition positive, $y_{i[m]}$ is also positive. This implies that for all three distributions

$$Var(\sum_i \frac{y_{ij}y_{ik}}{y_{i\cdot}}) \leq \sum_i E(\frac{y_{ij}^2 y_{ik}^2}{y_{i\cdot}^2}) \leq \sum_i E(\frac{y_{ij}^2 y_{ik}^2}{y_{i[m]}^2}) \leq \sum_i E(y_{ij}^2) = o(n^2).$$

Therefore, by the Weak Law of Large Numbers and using a Taylor's series expansion,

$$\sum_i \frac{y_{ij}y_{ik}}{y_{i\cdot}y_{\cdot j}} \rightarrow \frac{1}{\sum_i E(y_{ij})} \sum_i E(\frac{y_{ij}y_{ik}}{y_{i\cdot}}) = \sum_i \frac{E(y_{ij}y_{ik})}{E(y_{\cdot j})E(y_{i\cdot})} + \sum_i R_{ijk}, \quad (24)$$

where R_{ijk} is the remainder term. It follows that the j th element of \mathbf{v} satisfies

$$\lambda v_j = \frac{1}{E(y_{.j})} \sum_k v_k \sum_i \frac{Var(y_{ij}) \delta_{jk} + \mu_{ij} \mu_{ik}}{E(y_{i.})} + \sum_k v_k \sum_i R_{ijk}. \quad (25)$$

The first term on the right hand side of (25) equals

$$\frac{1}{\mu_{.j}} [v_j \sum_i \frac{\sigma_e^2}{\mu_{i.}} + \sum_i \frac{1}{\mu_{i.}} (a_j - b_j(x_i - u_j)^2) \sum_k v_k (a_k - b_k(x_i - u_k)^2)]$$

for the Gaussian case;

$$\frac{1}{\mu_{.j}} [\sum_i \frac{1}{\mu_{i.}} \exp(a_j - b_j(x_i - u_j)^2)] [v_j + \sum_k v_k \exp(a_k - b_k(x_i - u_k)^2)]$$

for the Poisson case; and

$$\frac{1}{\mu_{.j}} [v_j \sum_i \frac{1}{\mu_{i.}} r_{ij} (1 + r_{ij})^{-2} + \sum_i \frac{1}{\mu_{i.}} (1 + r_{ij})^{-1} \sum_k v_k (1 + r_{ik})^{-1}]$$

for the Bernoulli case, with $\mu_{.j} = \sum_i \mu_{ij}$, $\mu_{i.} = \sum_j \mu_{ij}$, and μ_{ij} defined according to the three

distributions respectively. The above three expressions are all nonlinear functions of u_j , which suggest that if the first order Taylor's series approximation were exact v_j cannot be linear in u_j and likewise $\hat{\mathbf{v}}$ cannot be a consistent estimator of \mathbf{u} .

The hypothesis that $\hat{\mathbf{v}}$ is inconsistent is now verified by numerically calculating \mathbf{v} by assuming that $x_i \sim \text{iid } N(0,1)$. According to (24), Σ has (j,k) th element

$$\frac{n}{\sum_i E(y_{ij})} \frac{1}{n} \sum_i E\left(\frac{y_{ij} y_{ik}}{y_{i.}}\right).$$

The first expectation can be computed using the same methods as described for the numerical calculations of the PCA asymptotic solutions, while the second expectation can be calculated using Monte Carlo integration. The same parameter inputs for calculating the PCA solutions in model (2) are used here again. Gaussian and Uniform random variables were generated using built-in functions in GAUSS, and Poisson and Bernoulli random variables were simulated using the inverse transform method. As shown in Figures 2a, 2b and 2c, the residuals $v_j - u_j$ are nonlinear functions of u_j , indicating that the CA estimator $\hat{\mathbf{v}}$ is inconsistent for model (2).

6. Simulations

Although the PCA and CA eigenvectors are generally inconsistent, applied researchers may be less concerned with their asymptotic behavior than with their finite sample performances. Therefore, simulations were conducted to compare PCA, CA, and ML under different parameter configurations for finite samples. In accordance with the two examples in Section 7, we concentrated on the Poisson and Bernoulli cases for model (2) and set $m=15$ and $n=25$. For parameter inputs, we let $a_j \sim \text{iid } U(0.5, 2.0)$, $b_j \sim \text{iid } U(1, 4)$, where $U(p, q)$ is a discrete Uniform distribution with equally spaced elements between (p, q) , and vary \mathbf{u} according to $E(0, 0.5)$, $E(0, 1.0)$, $E(-0.25, 0.25)$, or $E(-0.5, 0.5)$, and \mathbf{x} according to $E(-0.8, 0.8)$ or $E(-1.2, 1.2)$. The results are summarized in terms of the correlation between \mathbf{u} and its estimates, and each entry in Table 1 represents the mean of 1000 replications. Overall, the correlations are highest when the u_j 's have large variance relative to the variance of the x_i 's; e.g. $\mathbf{u}=E(-0.5, 0.5)$ or $E(0, 1.0)$. In contrast, when $\mathbf{u}=E(-0.25, 0.25)$ or $E(0, 0.5)$ the u_j 's have small variance relative to the x_i 's and the response curves are clustered together, resulting in weaker correlations. Changing the location of the u_j 's seems to have negligible effect on the CA estimates. But for PCA, the behavior of its first and second eigenvectors switched when $\mathbf{u}=E(-0.5, 0.5)$. In particular, the first PCA eigenvector has higher correlations when $\mathbf{x}=E(-0.8, 0.8)$ but the second PCA eigenvector has higher correlations when $\mathbf{x}=E(-1.2, 1.2)$. This can be compared to the Corollary in Section 5.1, which showed that in certain situations the second eigenvector might be the preferred estimator. A reviewer has also pointed out that the first PCA eigenvector usually represents the “size” dimension while the second PCA eigenvector measures “shape”, which is what CA analyses.

In terms of the comparative performance between CA and PCA, the CA estimates usually have stronger correlations. However, interestingly, PCA can also produce reasonable estimates, and the PCA second eigenvector did better in situations when the first eigenvector performed poorly. Their correlations are slightly weaker compared to those for the CA estimates, but for the Bernoulli case they can be comparable or stronger when the u_j 's have

small variance (e.g. when $\mathbf{x}=E(-1.2,1.2)$ and $\mathbf{u}=E(-0.25,0.25)$ or $E(0,0.5)$). ML estimates have the strongest correlations, although not substantially better than those from CA. Consequently, applied researchers may favor CA since ML computation for multi-parameter models is highly intensive and establishing convergence can be difficult due to the many local maxima (Mantel and Myers 1971, Dorsey and Mayer 1995). In addition, for the Bernoulli case the correlations are only between 0.5 and 0.7 when the u_j 's have small variance, suggesting the influence of small sample biases in the ML estimates.

7. Examples

Consider again the hunting spider example mentioned earlier (Ter Braak 1985, 1986). CA was used in the original analysis and a plot of the first and second order solutions showed a distinct arch pattern. Subsequently, only the first order CA row and column scores were used to describe the sites and species respectively. A Poisson regression of the species counts on the row scores suggested that model (2) provided an adequate fit. The row scores were also separately regressed on additional environmental data collected at the sites, and the results suggest that the row scores represent a composite latent gradient of soil moisture and openness of habitat. The conclusion was that wolfspiders required open habitats for hunting and also moisture to prevent desiccation, with each species balancing between these requirements and forming its specific niche along the environmental gradient. Our purpose here is to evaluate how well CA and PCA approximate the ordering of the species optima, and thus we focus on the column scores as estimates of the u_j 's rather than on the row scores, which are treated as nuisance parameters. The asymptotic results in Section 5 indicate that CA and PCA yield biased estimates with the quadratic Poisson model in (2), but the original analysis still suggested that there was a high correlation between the CA scores and the ML estimates. Indeed, the Pearson correlation between the ML estimates of u_j and the CA column scores is 0.97. This agrees with the simulation results, showing that CA performs well for the Poisson case and especially when the species optima are relatively spread out as seen from

plotting the predicted response curves from the Poisson regression analysis. The simulations also predict that with dispersed species optima the first PCA eigenvector will usually do better than the second PCA eigenvector, although still inferior to the CA solution. Likewise, when PCA is applied to the covariance matrix according to (5), the Pearson correlations between the ML estimates and the first and second PCA eigenvectors are 0.74 and 0.31 respectively. Overall, the results suggest that the CA column scores provide a realistic ordering of where each hunting spider species will tend to dominate along the latent gradient of soil moisture and habitat openness.

The second example involves incidence data, which are based on an archaeological excavation of 24 prehistoric grave sites. The artifacts collected at each grave were classified into 17 types to arrange the graves into some chronological order (Ihm and Groenewoud 1984). It is hypothesized that as time progresses one artifact type disappears and another type appears. If this hypothesis holds then the incidence matrix can be rearranged into a Petrie matrix, after the archaeologist Flinders Petrie, which has a block of consecutive ones in every row. Furthermore, the first order CA scores provide the correct ordering for the rearrangement (Greenacre 1984). For this example, rearranging the rows and columns using the CA scores did not yield an exact Petrie matrix but a general clustering of ones is evident along the diagonal. The Petrie matrix hypothesis also implies that under model (2) the u_j 's will span the entire range of the x_i 's. Under this condition, the simulations suggest that CA is the best alternative, followed by the first PCA eigenvector, when compared to ML. This is confirmed by the high Pearson correlation of 0.97 between the ML estimates and the CA column scores, while the correlation with the first PCA eigenvector is only 0.74.

8. Discussion

This research highlights two main messages. On the one hand, it cautions against an ad hoc use of PCA and CA for estimating parameters in models (1) and (2) without formal consideration of the inferential issues involved. Under the scenario with m fixed and $n \uparrow \infty$,

PCA, CA, and ML cannot yield consistent estimators unless sufficient “moment” conditions are imposed on x_i . The results show that the PCA estimator, which is the ML estimator for the linear Gaussian model, is inconsistent when applied to the Poisson and Bernoulli cases in models (1) and (2). However, the first and/or the second PCA eigenvectors can be location and scale consistent estimators of \mathbf{u} in the quadratic Gaussian model when b_j is constant. In contrast, the CA estimator is inconsistent for all three distributions in model (2). The above asymptotic setting seems most relevant for the examples considered in this paper. However, in certain applications it may be plausible for both m and n to increase to infinity, and then consistent ML estimators can be obtained without requiring any conditions on x_i (Haberman 1977, Portnoy 1988). Another asymptotic setting, which is appropriate for Poisson counts, assumes that m and n are fixed but the total sample count approaches infinity. This is the standard asymptotic theory for contingency table analyses, and there has been considerable discussion in this area on applying CA and relating it to RC association models (Goodman 1986, 1991, Van der Heijden *et al.* 1989). The incidental parameters problem is a consequence of treating x_i as a latent fixed effect. When the n objects can be regarded as a random sample of some population, x_i can be treated as a random effect and inference becomes applicable to the larger population of objects. It is therefore useful for researchers to carefully consider the appropriateness of fixed versus random effects in their models. Nevertheless, the consistency results for PCA and CA in Section 5 are valid for both cases since they were established by treating x_i as fixed and also verified by letting $x_i \sim \text{iid } N(0,1)$. The current models assume independence between observations, but correlations may be induced especially as n increases. For example, in the hunting spider example correlation may be induced because of the particular site environment or there may be dependence between different species due to symbiosis or competition effects. Such non-independence can be modeled by assuming random effects for a_j and x_i , and methods used for analyzing generalized mixed effects models can be applied (McCulloch 1997). With correlated observations, we expect PCA and CA to be inconsistent similar to the results for independent

observations, although it can be shown that the PCA estimator is consistent for the linear and quadratic Gaussian case under a compound-symmetry error covariance structure.

The other message this research conveys is that these computationally simple PCA and CA eigenvector solutions can often provide realistic approximations. Although PCA and CA in general do not lead to good estimators in terms of asymptotics, they can be satisfactory for finite sample sizes if one requires only an approximate ordering of the u_j 's. Using the correlation as a location and scale invariant measure for comparison, the results show that CA performs well when the optima of the different response curves are spread out over a large range of the x_i 's. In ecology, this condition is known as the “species packing model” assumption (Gauch *et al.* 1974). PCA usually produces weaker correlations compared to CA. However, the second PCA eigenvector can perform substantially better than the first eigenvector, and for incidence data with tightly clustered response curves it may do similar or better than CA. Hill (1973) has also pointed out the advantage of PCA on incidence data, although he stated that “the reasons why the method is successful with such data have not been made clear”. For small samples, PCA and especially CA can be comparable to ML, and may be recommended because of their computational ease. It may also be worthwhile to examine in future studies how PCA and CA compare in finite samples with ML or restricted ML estimation for generalized linear mixed effect models. However, it should be pointed out that PCA and CA do not directly estimate a_j and t_j in model (2) nor do they produce standard errors for their estimates.

REFERENCES

- Andersen, E. B. (1980), *Discrete Statistical Models with Social Science Applications*, Amsterdam: North Holland Publishing.
- Anderson, T. W. and Ameniya, Y. (1988), "The Asymptotic Normal Distribution of Estimators in Factor Analysis Under General Conditions," *The Annals of Statistics*, 16, 759-771.
- Aptech Systems (1992), *Gauss Manual*, Version 3.0, Kent, WA: Author.
- Birnbaum, A. (1968), "Some Latent Trait Models and Their Use in Inferring an Examinee's Ability". In: *Statistical Theories of Mental Test Scores* (Ed. by F. M. Lord and M. R. Novick), pp.397-472. Reading, MA: Addison-Wesley.
- Cox, C. P. (1972), "On Estimating Relative Potency From Quadratic Log-dose Response Relationships," *Biometrics*, 28, 875-881.
- Dawkins, B. (1989), "Multivariate Analysis of National Track Records," *The American Statistician*, 43, 110-115.
- Dorsey, R. E. and Mayer, W. J. (1995), "Genetic Algorithms for Estimation Problems with Multiple Optima, Nondifferentiability, and Other Irregular Features," *Journal of Business and Economic Statistics*, 13, 53-66.
- Follmann, D. (1988), "Consistent Estimation in the Rasch Model Based on Nonparametric Margins," *Psychometrika*, 53, 553-562.
- Fuller, W. A. (1987), *Measurement Error Models*, New York: Wiley & Sons.
- Gauch, H. G., Chase, G. B. and Whittaker, R. H. (1974), "Ordinations of Vegetation Samples by Gaussian Species Distributions," *Ecology*, 55, 1382-1390.
- Gauch, H. G., Whittaker, R. H., and Wentworth, T. R. (1977), "A Comparative Study of Reciprocal Averaging and Other Ordination Techniques," *Journal of Ecology*, 65, 157-174.
- Gleser, L. J. (1981), "Estimation in a Multivariate 'Errors-in-variables' Regression Model, Large sample results," *The Annals of Statistics*, 9, 24-44.
- Goodall, D. W., and Johnson, R. W. (1982), "Non-linear Ordination in Several Dimensions. A Maximum Likelihood Approach," *Vegetatio*, 48, 197-208.

- Goodman, L. A. (1986), "Some Useful Extensions of the Usual Correspondence Analysis Approach and the Usual Log-linear Models Approach in the Analysis of Contingency Tables," *International Statistics Review*, 54, 243-309.
- Goodman, L. A. (1991), "Measures, Models, and Graphical Displays in the Analysis of Cross-classified Data," *Journal of the American Statistical Association*, 86, 1085-1111.
- Goodman, L. A. and Haberman S. J. (1990), "The Analysis of Nonadditivity in Two-way Analysis of Variance," *Journal of the American Statistical Association*, 85, 139-145.
- Greenacre, M. J. (1984), *Theory and Application of Correspondence Analysis*, London: Academic Press.
- Haberman, S. J. (1977), "Maximum Likelihood Estimation in Exponential Response Models," *The Annals of Statistics*, 5, 815-841.
- Hill, M. O. (1973), "Reciprocal Averaging: an Eigenvector Method of Ordination," *Journal of Ecology*, 61, 237-249.
- Hill, M. O. (1974), "Correspondence Analysis: a Neglected Multivariate Approach," *Applied Statistics*, 23, 340-354.
- Ihm, P., and Van Groenewoud, H. (1984), "Correspondence Analysis and Gaussian Ordination," *COMPSTAT lectures*, 3, 5-60.
- De Leeuw, J. and Verhelst, N. (1986), "Maximum Likelihood Estimation in Generalized Rasch Models," *Journal of Educational Statistics*, 11, 183-196.
- Kendall, D. G. (1971), "Seriation from Abundance Matrices". In: *Mathematics in the Archaeological and Historical Sciences* (Ed. by C. R. Hodson, D. G. Kendall, and P. Táutu), pp.215-252. Edinburgh: Edinburgh University Press.
- Kiefer, J. and Wolfowitz, J. (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Nuisance Parameters," *Annals of Mathematical Statistics*, 27, 887-906.
- Kooijman, S. A. L. M. (1977), "Species Abundance with Optimum Relations to Environmental Factors," *Annals of System Research*, 6, 123-138.
- Lancaster, H. O. (1957), "Some Properties of the Bivariate Normal Distribution Considered in the Form of a Contingency Table," *Biometrika*, 44, 289-292.
- McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models, 2nd Edition*, London: Chapman and Hall.

- McCulloch, C. E. (1997), "Maximum Likelihood Algorithms for Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 92, 162-170.
- Mantel, N. and Myers, M. (1971), "Problems of Convergence of Maximum Likelihood Iterative Procedures in Multiparameter Situations," *Journal of the American Statistical Association*, 66, 484-491.
- Neyman, J., and Scott, E. L. (1948), "Consistent Estimators Based on Partially Consistent Observations," *Econometrica*, 16, 1-32.
- Ortega, J. M. (1972), *Numerical Analysis: A Second Course*, New York: Academic Press.
- Portnoy, S. (1988), "Asymptotic Behavior of Likelihood Methods for Exponential Families When the Number of Parameters Tends to Infinity," *The Annals of Statistics*, 16, 356-366.
- Ter Braak, C. J. F. (1985), "Correspondence Analysis of Incidence and Abundance Data: Properties in Terms of a Unimodal Model," *Biometrics*, 41, 859-873.
- Ter Braak, C. J. F. (1986), "Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis," *Ecology*, 67, 1167-1179.
- Ter Braak, C. J. F. (1987), "Ordination". In: *Data Analysis in Community and Landscape Ecology* (Ed. by R. H. G. Jongman, C. J. F. Ter Braak, and O. F. R. Van Tongeren), pp.91-173. Wageningen: Pudoc.
- Thisted, R. A. (1988), *Elements of Statistical Computing*, London: Chapman and Hall.
- Van der Heijden, P. G. M., Falguerolles, A. D., and Leeuw, J. D. (1989), "A Combined Approach to Contingency Table Analysis Using Correspondence Analysis and Log-linear Analysis," *Applied Statistics*, 38, 249-292.
- Van der Heijden, P. G. M., Mooijaart, A., and Takane, Y. (1994), "Correspondence Analysis and Contingency Table Models". In: *Correspondence Analysis in the Social Sciences. Recent Developments and Applications* (Ed. by M. Greenacre and J. Blaisus), pp.79-111. New York: Academic Press.
- Wright, B. D. (1977), "Solving Measurement Problems With the Rasch Model," *Journal of Educational Measurement*, 14, 97-116.

Figure 1. Asymptotic Solutions of PCA Estimators for Poisson (solid line) and Bernoulli (dotted line) in Model 1

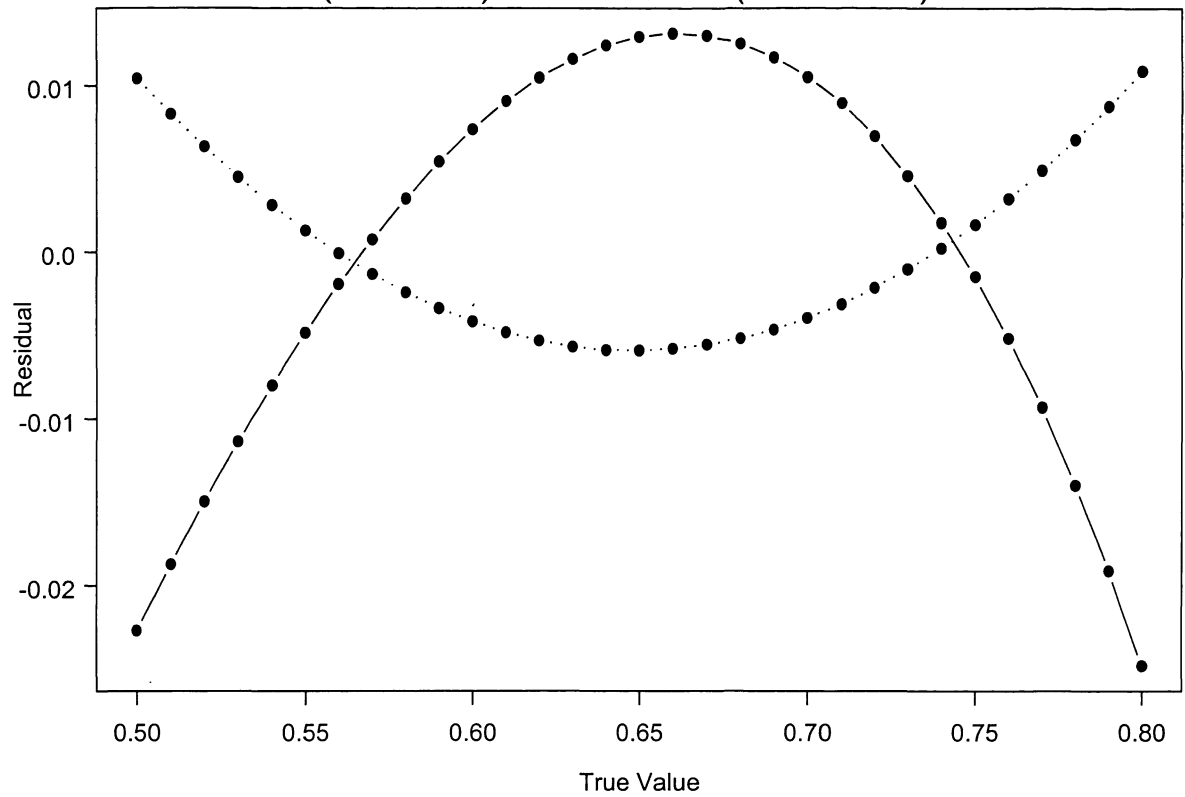


Figure 2a. Asymptotic Solutions of CA (solid line) and PCA (dotted line) Estimators for Gaussian Case in Model 2

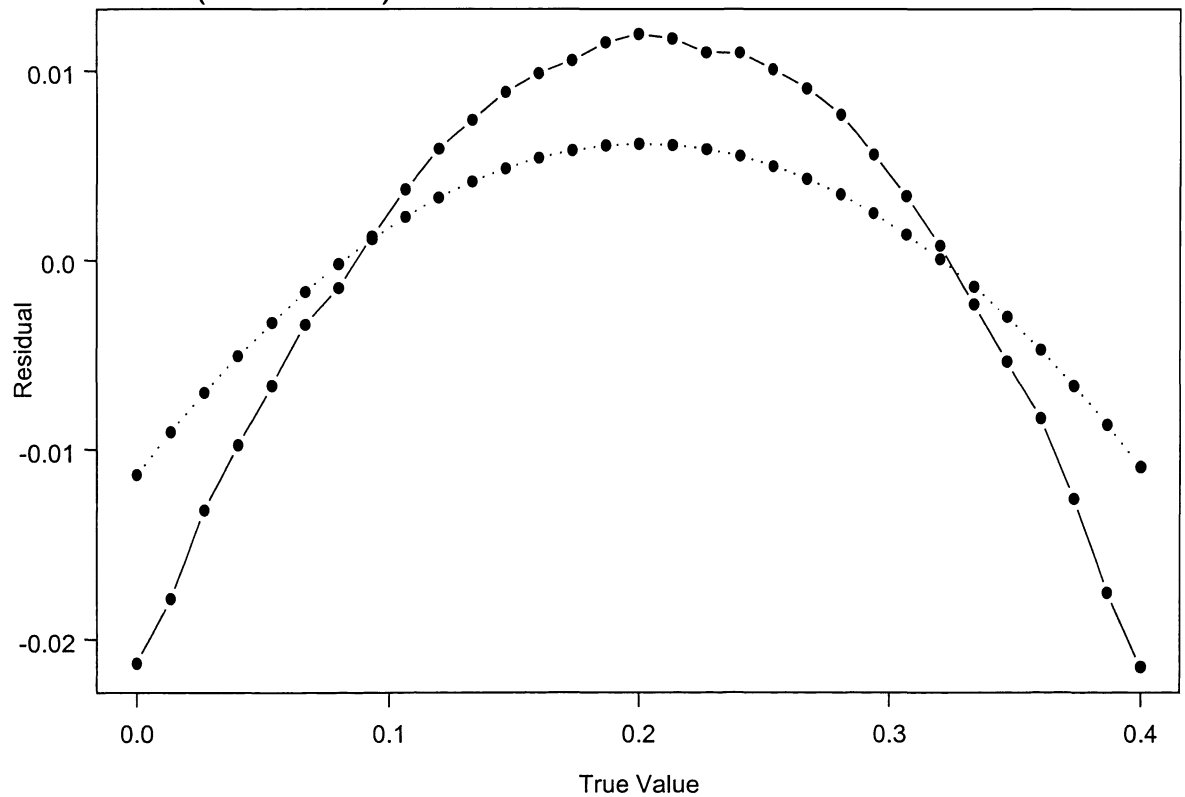


Figure 2b. Asymptotic Solutions of CA (solid line) and PCA (dotted line) Estimators for Poisson Case in Model 2

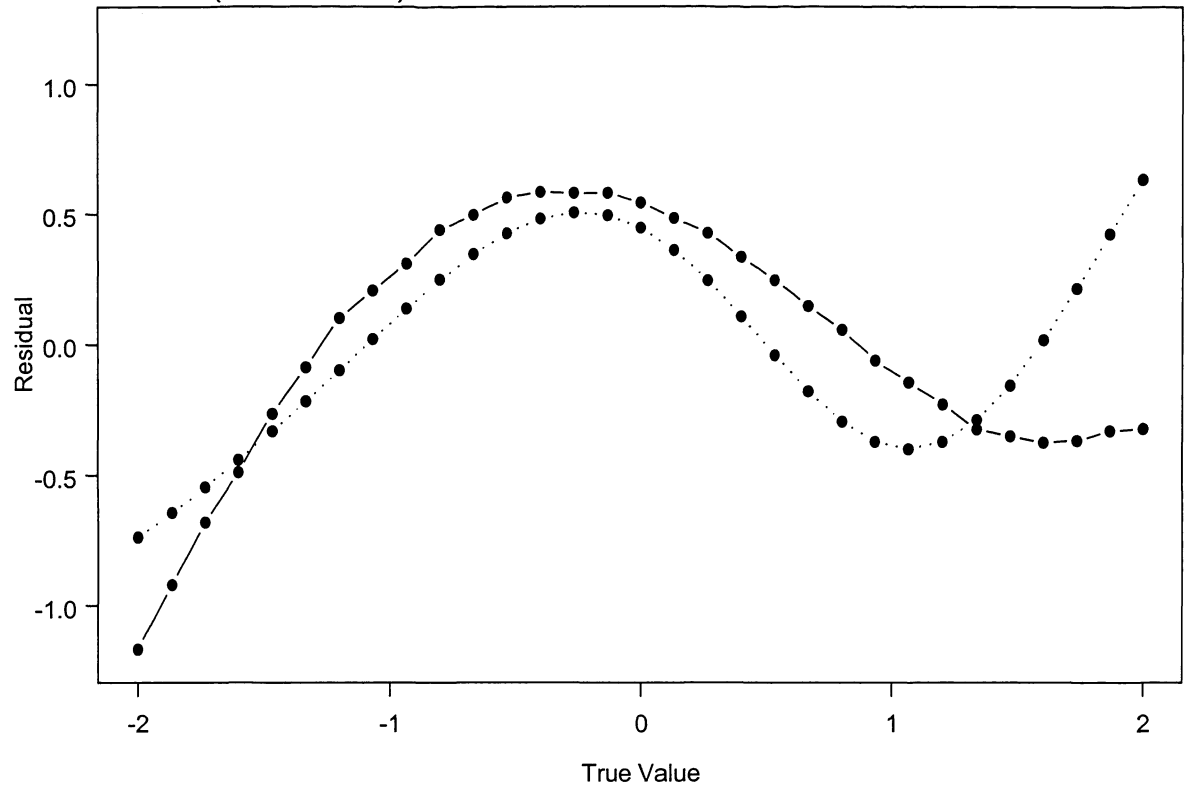


Figure 2c. Asymptotic Solutions of CA (solid line) and PCA (dotted line) Estimators for Bernoulli Case in Model 2

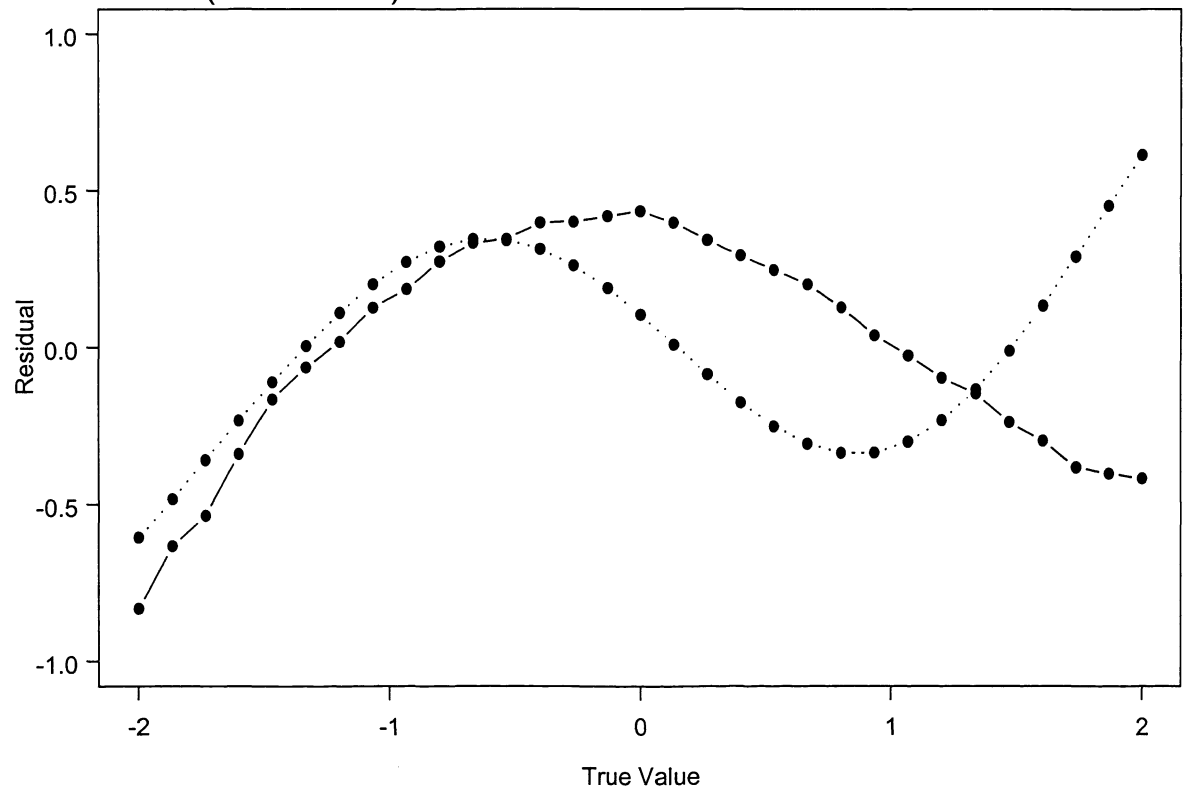


Table 1. Correlations between \mathbf{u} and its PCA, CA, and ML estimates for different parameter inputs according to model (2). Entries denote the mean of 1000 replications. For PCA, the entries are listed as $r1/r2$, where $r1$, $r2$ are the correlations of \mathbf{u} with its first and second order column scores respectively. We let $b_j \sim \text{iid } U(1,4)$ and $a_j \sim \text{iid } U(0.5,2.0)$, where $U(p,q)$ is a discrete Uniform distribution with equally spaced elements between (p,q) . $E(p,q)$ is a vector of equally spaced elements between (p,q) .

Poisson Case	\mathbf{u}	$\mathbf{x}=E(-0.8,0.8)$			$\mathbf{x}=E(-1.2,1.2)$		
		PCA	CA	ML	PCA	CA	ML
	$E(0,0.5)$	0.35/0.70	0.85	0.88	0.13/0.73	0.84	0.89
	$E(0,1.0)$	0.37/0.77	0.92	0.93	0.37/0.84	0.95	0.96
	$E(-0.25,0.25)$	0.24/0.76	0.87	0.89	0.13/0.74	0.84	0.89
	$E(-0.5,0.5)$	0.84/0.24	0.96	0.97	0.44/0.80	0.97	0.97

Bernoulli Case	\mathbf{u}	$\mathbf{x}=E(-0.8,0.8)$			$\mathbf{x}=E(-1.2,1.2)$		
		PCA	CA	ML	PCA	CA	ML
	$E(0,0.5)$	0.51/0.28	0.46	0.54	0.34/0.42	0.43	0.68
	$E(0,1.0)$	0.59/0.37	0.75	0.75	0.53/0.69	0.78	0.82
	$E(-0.25,0.25)$	0.37/0.33	0.41	0.57	0.25/0.45	0.42	0.70
	$E(-0.5,0.5)$	0.76/0.28	0.79	0.79	0.62/0.71	0.85	0.88